Supplementary material for: Comparative tests of the role of dewlap size in *Anolis* lizard speciation

Image compilation and data extraction

We compiled side-view photographs of adult male anoles with fully extended dewlaps. In most the dewlap was manually extended (typically with forceps), but some anoles were photographed with their dewlaps naturally extended. The two species without dewlaps were also photographed with the skin below the throat extended with forceps, allowing measurements comparable to dewlap area. Photos were taken by multiple investigators in field locations throughout the geographic range of anoles. Investigators confirmed species identities in the field, and provided collection information and snout-vent length (SVL) if available. In most cases a ruler, grid, or object of known size was in the photo for scale. We supplemented these photographs with photos of anoles with extended dewlaps from a field guide [1] and from the CaribHerp database (`www.caribherp.org`). If multiple photos were available for an individual we selected the one with the best resolution and the best lateral view of the head and dewlap. We measured only males, which are more frequently photographed and have both larger dewlaps on average and greater interspecific variation in dewlap size, although female dewlap size is also variable and warrants further study [2].

We opened each photo in the image analysis program ImageJ (`http://rsbweb.nih.gov/ij/`), and set the scale. For photos lacking a scale object we temporarily set the scale in units of the lizard's head length. As an operational measure of head length that was visible in all images, we measured the distance from the tip of the snout to the posterior margin of the ear. Dewlap area was measured by drawing a straight line from the anterior to posterior insertion points of the dewlap, then tracing the outer margin of the dewlap using the "polygon-select" tool (figure 1). If the dewlap was partly obscured by the lizard's foreleg or the researcher's finger or forceps, photos were only used if the obscured part of the dewlap was small enough for its position to easily be inferred. All image analysis was done by one investigator (TI).

To allow measurement of images without an absolute scale, we first carried out a regression of mean log head length against mean log SVL for species with SVL data for photographed individuals ($n = 94$, $\log \text{HL} = -3.54 + 0.999 \times \log \text{SVL}$; $R^2 = 0.93$). We then estimated head lengths for each

photograph lacking scale using species' published adult SVL [3, 4] and rescaled the measured dewlap area in proportion to the estimated head length. To ensure that incorporating data from photos without scale did not substantially affect our estimates dewlap measurements, we compared mean values calculated separately for species for which we had photos with and without scale. For 53 species with both direct measures and indirect estimates of head length, there were strong correlations between the two values for mean log head length ($r = 0.96$) and mean log dewlap area ($r = 0.93$).

We used our measure of head length to represent lizard body size, to allow inclusion of individuals without SVL data. We acknowledge that species vary in relative head length [5] and that this will have some effect on our estimates of relative dewlap size, but this should not cause any systematic bias in our later analyses. To measure dewlap area relative to lizard size, we carried out a phylogenetic regression of species mean log dewlap area against species mean log head length, using the R function 'gls' and the MCC tree. We simultaneously estimated the appropriate level of phylogenetic signal in the residuals by optimising the λ parameter (hereafter 'λ_{sig}' to distinguish it from the speciation rate λ) [6]. λ_{sig} varies from zero to one or slightly larger than one. Values near one indicate strong phylogenetic signal as expected under a model of random walk Brownian motion. As λ_{sig} approaches zero, phylogenetic signal is progressively weaker until trait similarity is effectively independent of phylogenetic relatedness. We compared the fit with the estimated λ_{sig} to fits with λ_{sig} fixed at one or zero.

Using the the coefficients from the phylogenetic regression (log dewlap area $= -0.51 + 1.99\times$ log head length), we first calculated residual log dewlap area for each individual lizard, then calculated the mean relative dewlap area for each species. This allowed us to estimate intraspecific variability as the standard deviation and standard error of the mean for each species, for use in subsequent comparative analyses. For species represented by a single individual, we calculated the mean intraspecific standard deviation across species with multiple individuals, and substituted this value as an estimate of both the standard deviation and standard error.

We assessed measurement error associated with photography, digitisation and size-correction using individual lizards that were photographed twice. We first used a set of 10 anoles of different, arbitrarily selected species to assess measurement error relative to interspecific variation, and extracted measurements from a second photograph of each individual. We calculated the intraclass correlation coefficient [7] for log dewlap area, log head length, and relative dewlap area, and found extremely high repeatability (all intra-class correlation coefficients > 0.99). We repeated this exercise

with a set of 10 individual *A. cybotes*, to assess measurement error relative to intraspecific variation, and found that the repeatability was slightly lower but still high for all traits (intra-class correlation coefficients > 0.94).

The use of images from CaribHerp and field guides might affect our data quality if these differ consistently from photos taken in the field. To assess this possible influence, we calculated species mean residual dewlap area using only individuals taken from CaribHerp or a field guide, and again using only individuals photographed in the field. 55 species were represented in both field photos and in CaribHerp (the four species taken from a field guide were not represented in field photos). We calculated the Pearson's correlation between species means from field photos and species means from CaribHerp, and found that the two estimates were strongly correlated ($r = 0.90$).

Implementation and evaluation of the 'multi-ψ' model

The ψ model was introduced as a more intuitive representation of the contribution of gradual and speciational trait evolution on a phylogeny [8]. Gradual evolution is modeled as unbounded Brownian Motion (BM) with rate parameter σ_a^2 (for 'anagenetic' evolution), with the expected phenotypic divergence between species proportional to the amount of time elapsed since speciation. In contrast, speciational evolution is modeled as effectively instantaneous step change in the traits of each of the two daughter species, with each new species' trait change drawn from Gaussian distribution with variance σ_c^2 (for 'cladogenetic' evolution) [9]. As gradual and speciational evolution have different effects on the expected difference between pairs of species' traits (related to the amount of time and the number of speciation events since their common ancestor, respectively), it is possible to estimate the contributions of the two modes of trait evolution.

Parameter estimation employs a branch length transformation that varies the contribution of gradual and speciational change. With an original branch length b_i in units of time (generally m.y.), the transformed branch length is $b'_i = \sigma_c^2\, S(i) + \sigma_a^2\, b_i$. $S(i)$ is the number of speciation events on this branch, including the origin of the branch and any additional speciation events (see below). A simple reparameterisation of this model defines the total rate of evolution as $\sigma_t^2 = \sigma_a^2 + 2\lambda\sigma_c^2$, putting both variance terms on the same timescale. The '2' reflects the assumption that both daughter species diverge at each speciation event rather than only one - this was absent in the original description of the ψ model [8], though either assumption is justifiable. The parameter ψ is then defined as the proportion of σ_t^2) that is attributable to speciational evolution ($\psi = 2\lambda\sigma_c^2/\sigma_t^2$). Under the new

parameterisation, branches are transformed as:

$$b_i' = \sigma_t^2 \left(\frac{\psi}{2\lambda} S(i) + (1 - \psi) b_i \right), \tag{1}$$

with ψ bounded between 0 (gradual evolution by BM) and 1 (fully speciational evolution). The likelihood is calculated as the multivariate normal distribution of species traits, accounting for their phylogenetic relationships and the model parameters, and the maximum likelihood parameter values are obtained through optimisation [10, 11].

Inferring the contribution of speciational evolution is straightforward if all speciation events are represented in the tree, but this is only true of a fully-sampled clade with no extinction. We can retain the positions of any 'missing' nodes that are deleted when species lacking trait data are pruned from the tree, but more challenging are nodes 'hidden' because of extinct or unsampled lineages. We use estimated speciation and extinction rates λ and μ [12] to calculate the expected number of 'hidden' speciation events S_h on each branch [9]. We then sampled multiple realisations of S_h on each branch from a Poisson distribution, and repeated analyses on each realisation.

In this manuscript we extend the ψ model to allow different values of ψ for different parts of the phylogeny, similar to methods that fit multiple Brownian rate parameters [13, 14]. This allows for the possibility that only certain lineages undergo speciational evolution, potentially indicating different modes of speciation. We assume that the total rate of evolution σ_t^2 is constant across the tree, and that each branch can be placed in one state defined by geography or some other characteristic (consistent with the 'noncensored' approach used by O'Meara [13]. As with related multi-state models that use stochastic character mapped trees, it should be possible to allow state transitions to occur midway along a branch as well as at the origin of the branch. Currently this has not been implemented because the procedure of sampling hidden speciation events assigns events to a branch but not to a specific position on that branch, complicating the assignment of discrete speciation events to discrete states if states change along branches.

The ψ and multi-ψ models are newly implemented in two analytic frameworks. The first uses a similar analytic method to the 'Brownie' method for fitting multple BM rate models, and is modified from the 'phytools' [15] function 'brownie.lite'. This method is used for the analyses presented in the manuscript, and the function and related code are included as an appendix. The second implementation uses the framework of the R package motmot [10, 11] and is available at

https://github.com/ghthomas/motmot/tree/motmot_psi/R. This modeling framework takes advantage of fast likelihood calculations using phylogenetically independent contrasts, and leads to estimates of ψ that are highly similar to the first implementation, though the incorporation of measurement error and intraspecific variation is slightly different.

We analysed simulated data to test the ability of the multi-ψ model to recover true variation in the relative importance of speciational evolution under a range of conditions. We began by simulating birth-death trees using the function 'sim.bdtree' in geiger [16], half with no extinction ($\lambda = 0.5$, $\mu = 0$) and half with extinction 50% of the speciation rate ($\lambda = 0.5$, $\mu = 0.25$). We simulated until an extant diversity of 32, 64, 128 or 256 taxa was reached, and kept both the complete tree with extinct lineages and the 'reconstructed' tree with only extant taxa. For each tree, we identified one node at which the 'state' would change, selecting the node whose descendants comprised as close as possible to 50% of the extant taxa in the tree.

We selected fifteen combinations of two ψ values, including all unique combinations of $\psi = 0$, 0.25, 0.5, 0.75, and 1. Note that this parameter set includes BM ($\psi_1 = 0, \psi 2 = 0$) as well as four versions of the single-ψ model ($\psi_1 = \psi_2 = 0.25$, 0.5, 0.75 or 1) and ten versions of the multi-ψ model. We simulated trait evolution with the given ψ values, with 50 replicates of each combination of ψ values, tree size, and extinction rate. We fit the BM, multi-BM, ψ, and multi-ψ models to each simulated data set, and evaluated the accuracy of estimation of ψ values (figure S2) and the ability of the method to correctly identify the true model using AIC_c (figure S3). Estimates of ψ were variable but generally unbiased and more accurate for larger trees. In general, the multi-ψ model could generally be recovered if the tree was large and/or the difference between the two true ψ values was more substantial.

Results using absolute dewlap area

The analyses presented in the main text focus on dewlap size measured relative to the size of the lizard's body (specifically by measuring each dewlap area relative to predicted values from a log-log regression of species mean dewlap area on species mean head length). This measure has the advantage of avoiding inference of a role of dewlap size in either trait-dependent speciation or speciational trait evolution when the fit instead reflects a role of body size. Across species, absolute dewlap size was slightly more strongly correlated with our measure of body size (head length; $r = 0.74$) than it was with relative dewlap size ($r = 0.64$), so any results for absolute dewlap size may

represent the influence of overall body size (the ecological importance of which is well-established in anoles) in addition to the influence of the dewlap per se. We also feel that when dewlaps are used in social interactions between individuals of the same or closely related species, the size of the signal relative to the signaller is more likely to contain information than the absolute size of the signal (as a signal on an individual of a much larger or much smaller species is more likely to occur in a predator-prey context).

However, as the absolute size of the dewlap may be the more important determinant of its visibility (in conjunction with colour and contrast), we repeated our main analyses using species mean log-transformed absolute dewlap area as the focal trait. We first ran QuaSSE and split-QuaSSE analyses using log-transformed absolute dewlap area and the MCC tree, with other settings as described for residual dewlap area. Including a linear relationship between speciation rate and absolute DA did not improve either the standard or split model results, and the best fit was the split-constant model with separate speciation rates for island and mainland lineages but no relationship with absolute dewlap area (next best model split-linear; ΔAIC = 1.35). We also fit the six models of trait evolution to absolute dewlap size using the MCC tree. As was the case for relative dewlap size, the multi-ψ model had the best fit ($\psi_{\text{island}} = 0, \psi_{\text{mainland}} = 0.74$, Akaike weight = 0.58). The single-ψ model did not perform as well as it did for relative dewlap size ($\psi = 0.14, \Delta\text{AIC} = 7.65$, Akaike weight = 0.01), and the single-optimum OU model was the second-ranked model (ΔAIC = 2.64).

References

[1] Köhler G. Reptiles of Central America. Offenbach: Herpeton; 2008.

[2] Harrison A, Poe S. Evolution of an ornament, the dewlap, in females of the lizard genus *Anolis*. Biol J Linn Soc. 2012;106:191–201.

[3] Meiri S. Evolution and ecology of lizard body sizes. Global Ecol Biogeogr. 2008;17:724–734.

[4] Mahler DL, Revell LJ, Glor RE, Losos JB. Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. Evolution. 2010;64(9):2731–2745.

[5] Sanger TJ, Sherratt E, McGlothlin JW, Brodie ED III. Convergent evolution of sexual dimorphism in skull shape using distinct developmental strategies. Evolution. 2013;67:2180–2193.

[6] Freckleton R, Harvey P, Pagel M. Phylogenetic analysis and comparative data: a test and review of evidence. Am Nat. 2002;160:712–726.

[7] Wolak ME, Fairbairn DJ, Paulsen YR. Guidelines for estimating repeatability. Methods Ecol Evol. 2012;3:129–137.

[8] Ingram T. Speciation along a depth gradient in a marine adaptive radiation. Proc R Soc Lond B. 2011;278(1705):613–618.

[9] Bokma F. Detection of "punctuated equilibrium" by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. Evolution. 2008;62:2718–2726.

[10] Freckleton RP. Fast likelihood calculations for comparative analyses. Methods Ecol Evol. 2012;3:940–947.

[11] Thomas GH, Freckleton RP. MOTMOT: models of trait macroevolution on trees. Methods Ecol Evol. 2012;3:145–151.

[12] Rabosky DL. LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. Evol Bioinform Online. 2006;2:247.

[13] O'Meara BC, Ané C, Sanderson MJ, Wainwright PC. Testing for different rates of continuous trait evolution using likelihood. Evolution. 2006;60:922–933.

[14] Thomas GH, Freckleton RP, Székely T. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proc R Soc Lond B. 2006;273(1594):1619–1624.

[15] Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 2012;3:217–223.

[16] Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. GEIGER: investigating evolutionary radiations. Bioinformatics. 2008;24:129–131.

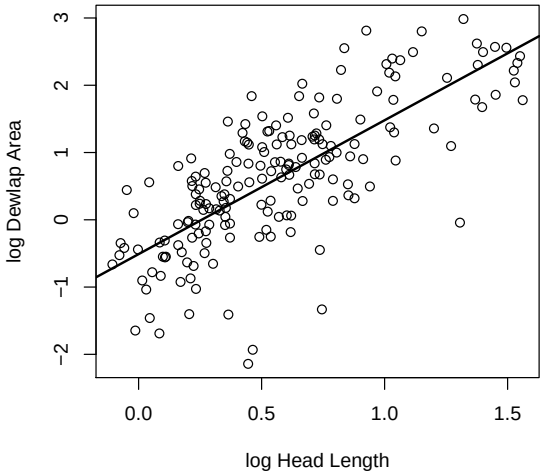

Figure S1. Relationship between species mean head length (measured in cm, log transformed) and species mean dewlap area (measured in cm^2, log transformed), with the fitted phylogenetic regression line used to calculate residual dewlap area.

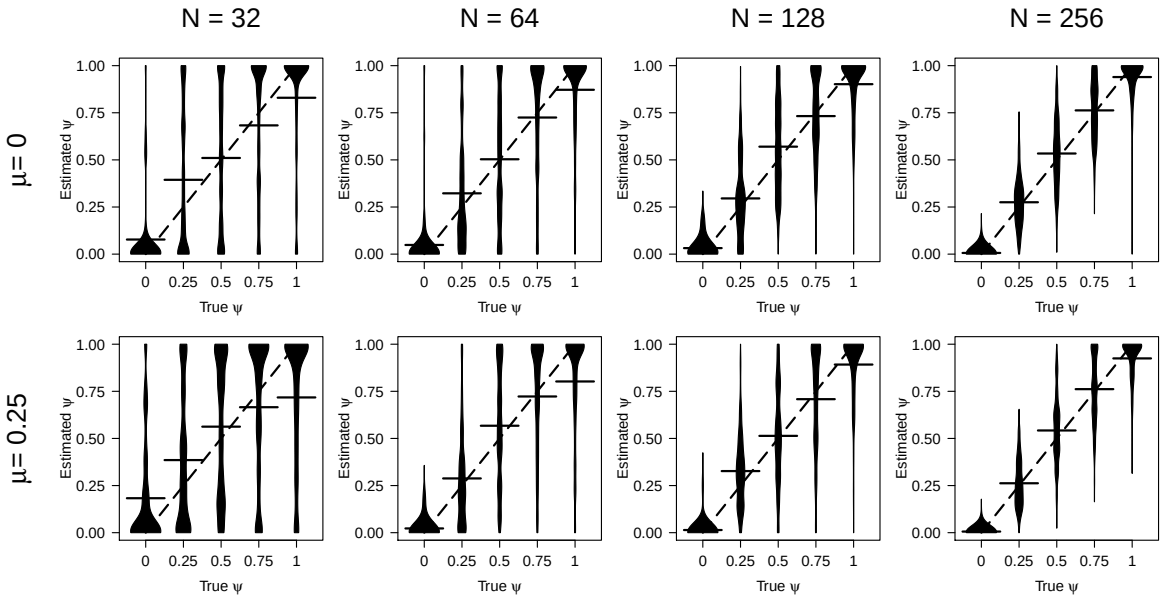

Figure S2. Accuracy of estimating the true ψ value for one of two states using the multi-ψ model (results were highly similar regardless of which state was analysed). 'Beanplots' visualise the kernel density estimates of the distribution of values of ψ estimated for each true value of ψ.

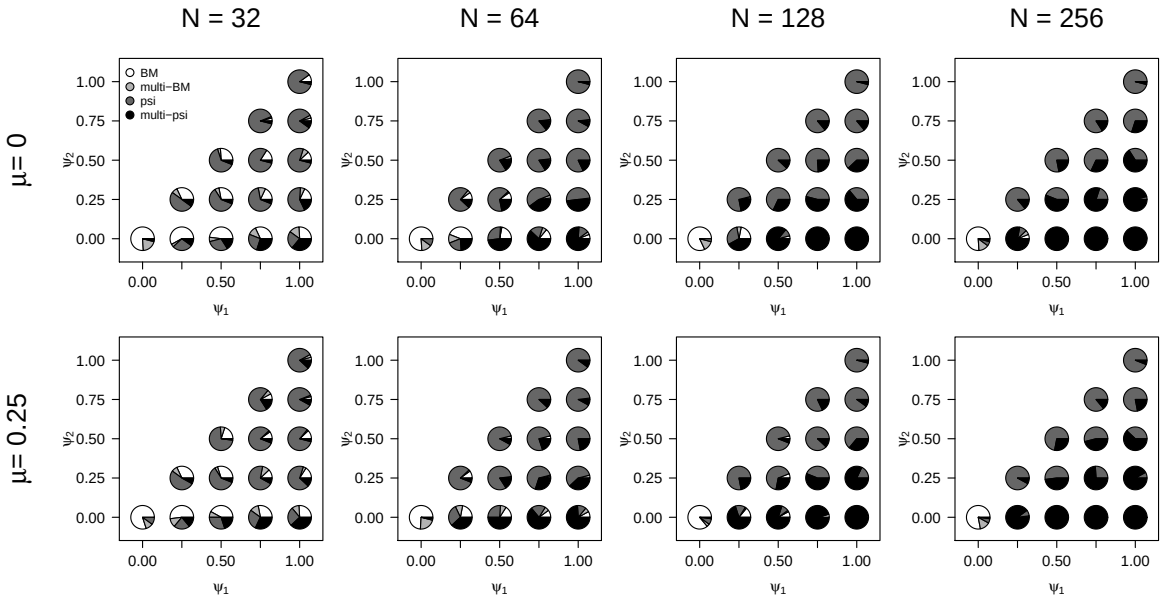

Figure S3. Frequency with which the true model was favoured by AIC_c for a range of simulated data sets. Each pie shows the frequency with which each of four models had the lowest AIC_c. Note that the true generating model was BM for the lower left pie in each panel, single-ψ for the diagonal ($\psi_1 = \psi_2$, and multi-ψ for the off-diagonal pies.